



# Combined use of algorithms for peak picking, peak tracking and retention modelling to optimize the chromatographic conditions for liquid chromatography–mass spectrometry analysis of fluocinolone acetonide and its degradation products

Mattias J. Fredriksson<sup>a</sup>, Patrik Petersson<sup>b</sup>, Bengt-Olof Axelsson<sup>b</sup>, Dan Bylund<sup>a,\*</sup>

<sup>a</sup> Department of Natural Sciences, Engineering and Mathematics, Mid Sweden University, SE-851 70 Sundsvall, Sweden

<sup>b</sup> Astra Zeneca R&D Lund, SE-221 87 Lund, Sweden

## ARTICLE INFO

### Article history:

Received 28 February 2011

Received in revised form 13 July 2011

Accepted 23 July 2011

Available online 10 August 2011

### Keywords:

Liquid chromatography

Mass spectrometry

Method development

Peak tracking

## ABSTRACT

A strategy for rapid optimization of liquid chromatography column temperature and gradient shape is presented. The optimization as such is based on the well established retention and peak width models implemented in software like e.g. DryLab and LC simulator. The novel part of the strategy is a highly automated processing algorithm for detection and tracking of chromatographic peaks in noisy liquid chromatography–mass spectrometry (LC–MS) data. The strategy is presented and visualized by the optimization of the separation of two degradants present in ultraviolet (UV) exposed fluocinolone acetonide. It should be stressed, however, that it can be utilized for LC–MS analysis of any sample and application where several runs are conducted on the same sample. In the application presented, 30 components that were difficult or impossible to detect in the UV data could be automatically detected and tracked in the MS data by using the proposed strategy. The number of correctly tracked components was above 95%. Using the parameters from the reconstructed data sets to the model gave good agreement between predicted and observed retention times at optimal conditions. The area of the smallest tracked component was estimated to 0.08% compared to the main component, a level relevant for the characterization of impurities in the pharmaceutical industry.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

The development of liquid chromatography (LC) methods for the determination of impurities and degradation products in drug products is a central task within the pharmaceutical industry. All impurities and degradation products present at levels above 0.1% of the active substance must be quantified, identified and qualified. The most commonly used techniques are LC combined with ultraviolet (UV)–detection for quantitative analysis and LC combined with mass spectrometry (MS) for identification [1]. MS detection is also frequently used during the development of LC–UV methods to facilitate peak tracking, i.e. identify how peaks move as the chromatographic conditions change.

Several efforts have been made to reduce the time needed for LC analysis as well as the LC method development itself [2,3]. Faster separations can be achieved by improvements in the instrumentation such as the use of monolithic columns or systems employing columns packed with sub 2 μm porous or 2.7 μm superficially

porous particles in combination with elevated temperature and back pressures (600–1200 bar). A reduction of the time needed for method development has also been achieved by the introduction of automated column and solvent selection valves that reduces user intervention. A number of method development strategies have been developed to decrease the total number of trials needed to find optimal conditions for the determination of degradation products and/or impurities [4–13].

Optimization/prediction software has been developed that, based on a small number of experiments, can guide the analyst towards optimal conditions. DryLab [14,15], LC simulator/Autochrom [16], ChromSword [10,17], Osiris [18] and LabExpert [9] are examples on data programs that have been developed for these tasks and in some cases even take control over the instrument and is capable to make decisions on what chromatographic parameters to try next to reach optimal conditions. Many of the optimization programs use simple but efficient relations between variables and responses that require a minimum number of analytical runs. Similar strategies based upon experimental design and multivariate evaluation has also been described in the literature, for example by Moberg et al. [19–21] and Popovic et al. [4]. Also, a combination of these approaches has been reported [6,7].

\* Corresponding author. Tel.: +46 60 148909; fax: +46 60 148820.  
E-mail address: [dan.bylund@miun.se](mailto:dan.bylund@miun.se) (D. Bylund).

The development part of AstraZeneca has adopted a three-step strategy which appears to be common in the pharmaceutical industry. An initial scouting step where a suitable pH value and buffer are selected based on the structure and  $pK_a$  values of the drug substance, by prior knowledge or by screening mobile phases with different pH. The purpose with this step is to find buffers that produce a good peak shape as well as sufficient retention. A second step comprises a generic screening of different combinations of columns [22], organic modifiers and buffers with different pH values which has been selected to maximize selectivity differences. The purpose with this step is to determine which components are present in the samples as well as to find a good starting point for the subsequent optimization. A third and final step is the actual optimization in which the temperature of the column and the shape of the gradient are simultaneously optimized. This optimization is based on well known retention time and peak width models [23], which have been implemented in the commercial software like DryLab [14], LC simulator/Autochrom [16], ChromSword [10] and Osiris [18].

These software and most method development strategies require peak tracking. In most cases a challenging and time consuming part of the method development process. Unfortunately, not much has been reported on how the time needed for peak tracking can be reduced. Diode array detection (DAD) is often of limited use since impurities and degradants related to the active component typically are present at low level where the quality of the UV spectrum is poor. An additional complication is that related impurities tend to have a spectrum very similar to the active component. Peak tracking based on UV peak area is often not reliable due to poor signal to noise ratio, which results in an uncertain determination. Another complication is that degradation may change peak areas over time [24]. MS detection is probably the most reliable tool for peak tracking. It is, however, often a challenge to locate chromatographic peaks at relevant levels in the relatively noisy LC–MS data matrices. Another problem is the large amount of data produced, which is quite time consuming to process manually. This is currently considered to be one of the largest bottlenecks in the method development process. Hence, there is a need for automated processing methods for detection and tracking of chromatographic peaks.

In this work such an automated processing method is described. Its application is exemplified by the optimization of a LC–MS method for the determination of two degradants present in fluocinolone acetonide after exposure to UV light [25,26]. The objective is to show how our previously developed processing methods for signal enhancement, peak picking [27] and peak tracking [28] can be combined to facilitate retention modelling and the optimization of chromatographic conditions.

## 2. Experimental

### 2.1. Sample preparation

A solution of fluocinolone acetonide (CAS 67-73-2, Sigma–Aldrich, St. Louis, MO) was prepared at  $1\text{ mg mL}^{-1}$  in 50/50 (v/v) acetonitrile/water. This solution was then exposed to the ambient light and temperature in a transparent glass vessel for 24 h to accelerate degradation of the steroid. The final solution was stored in darkness at  $-18^\circ\text{C}$  pending analysis.

### 2.2. Chemical analysis

The LC–DAD–MS system used comprised two Micro Series 2000 LC pumps (PerkinElmer, Inc., Wellesley, MA), an autoinjector (G1367A, Agilent Technologies, Waldbronn, Germany), a Kinetex C18 column ( $100\text{ mm} \times 2.1\text{ mm}$ ,  $2.6\ \mu\text{m}$ ) from Phenomenex, Inc. (Torrance, CA), a column oven (PerkinElmer), a DAD (G1315B,

Agilent) and an API 3000 mass spectrometer (AB Sciex, Concord, Ontario Canada). Gradient elution at various temperatures and time programs was applied with mobile phase A and B at a flow rate of  $0.4\text{ mL min}^{-1}$ , where A consisted of 5% and B of 95% acetonitrile (HPLC gradient grade, Sigma–Aldrich) in 0.1% (v/v) aqueous formic acid (analytical grade, Sigma–Aldrich). The injection volume was  $6\ \mu\text{L}$  and the DAD was operated within the range 210–600 nm. Positive electrospray ionization was applied with the mass spectrometer operated in full Q1 scan mode for  $m/z$  100–950 using default settings. The sampling frequencies for the detectors were adjusted to 20 (DAD) and 2.47 (MS) Hz, respectively. 2.47 Hz was the highest possible scan speed obtainable with acceptable signal on the current MS instrument. A higher frequency should be preferable for the column used, but as optimal performance is not reached in a conventional LC system (for the current set-up, reduced plate heights of  $h = 5.5$  were reached under isocratic conditions, which is about three times the optimal plate height of the column), 2.47 Hz gave satisfactory results. The dwell volume and dead time were determined to 0.89 mL and 1.03 min respectively. The sample was stored in darkness at  $5^\circ\text{C}$  in the auto injector during analysis to reduce the extent of further degradation.

### 2.3. Data analysis

Analyst 1.3 (build 4301) was used for data acquisition, and then the raw data was transferred to MATLAB format by use of the wiff-to-matlab add-on. In MATLAB 7.0.4.365 (R14), peak detection and peak tracking was performed according to the procedures described in our previous publications [27,28]. Thereby peaks in common for the separate runs could be identified and characterized by their retention times and widths. These peak lists were then, together with the chromatographic conditions, fed into the chromatography optimization software DryLab, where retention models were generated from which optimal gradients and temperatures for the separation task could be determined.

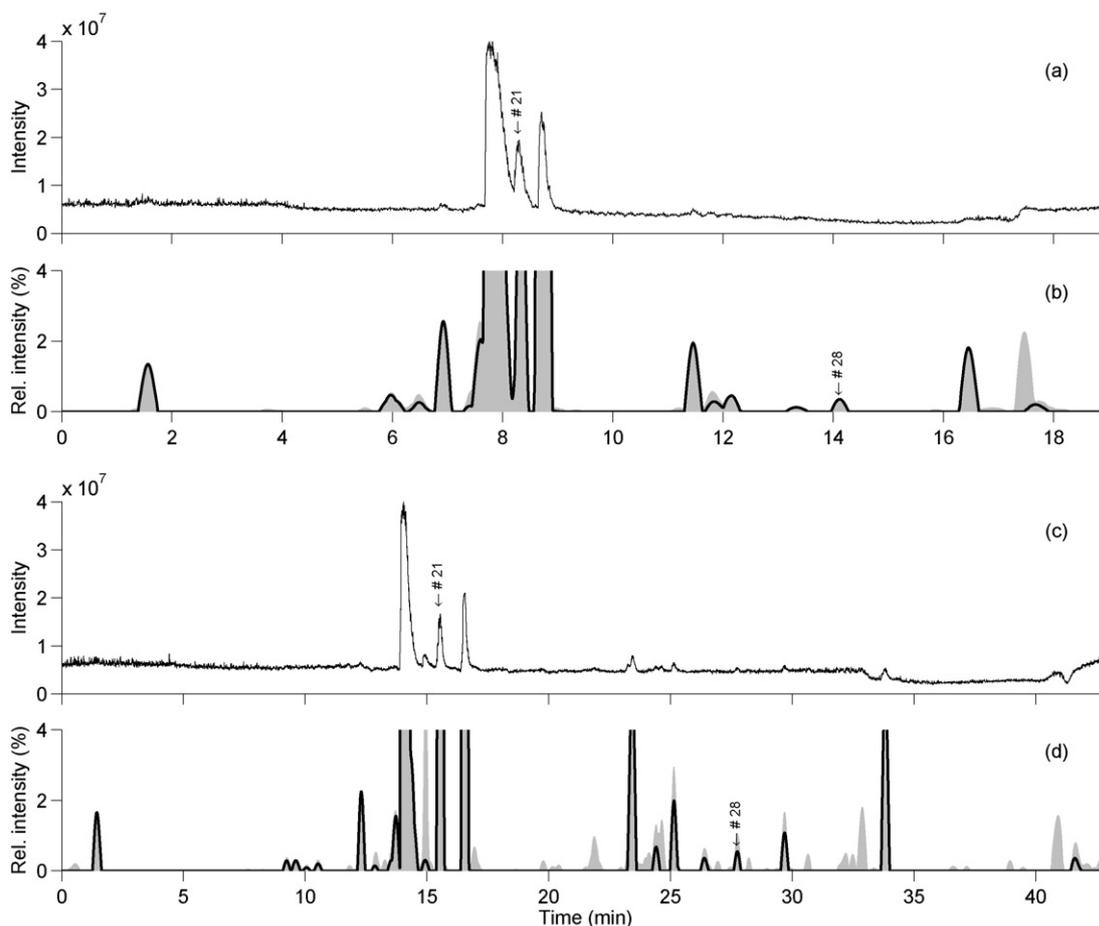
## 3. Results and discussion

The scope of this article is to demonstrate how our previously developed processing methods for signal enhancement, peak picking [27] and peak tracking [28] can be combined to facilitate retention modelling and the optimization of chromatographic conditions, i.e. the final step in a typical method development strategy. It should be stressed, however, that these processing methods are equally applicable for the detection of relevant impurities in the screening step.

### 3.1. The model data sets

In order to generate retention models required for optimization, six data sets were acquired at two different temperatures, and three different gradient slopes. Prior to processing, the data (Fig. 1) showed a major peak in the TIC with a dominating  $m/z$  of 454 corresponding to that of protonized fluocinolone acetonide. Also clearly visible are two peaks eluting after the main peak which are present at TIC areas of approximately 15 and 25%, respectively. In some of the data sets, peaks in the order of 1% compared to the main peak are distinguished in the TIC or BPC representations. In Fig. 1(a) and (c), the TICs for the steepest gradient acquired at the lowest temperature and for the shallowest gradient at the highest temperature are shown.

All six data sets were subjected to the previously developed peak detection algorithm. During this step, the chromatographic peaks are extracted from the noise and background in each mass channel. In the general case, not much attention is needed for the generated reconstructed data set since the process is fully automated and the



**Fig. 1.** (a) TIC of the data set sampled at 25 °C and steep gradient with a gradient time of 12 min and a range from 5 to 95% ACN. (b) The gray filled peaks constitute the sum of the automatically detected and estimated component profiles of the data set depicted in (a) and the solid line corresponds to the sum of the components that matches in all six model data sets truncated at 4% of max intensity. (c) TIC of the data set sampled at 50 °C and shallow gradient with a gradient time of 36 min (same range as in (a)). (d) Same as in (b), but corresponding to the dataset depicted in (c). The peak area of the smallest peaks visible in (d) corresponds to 0.08% of the main component.

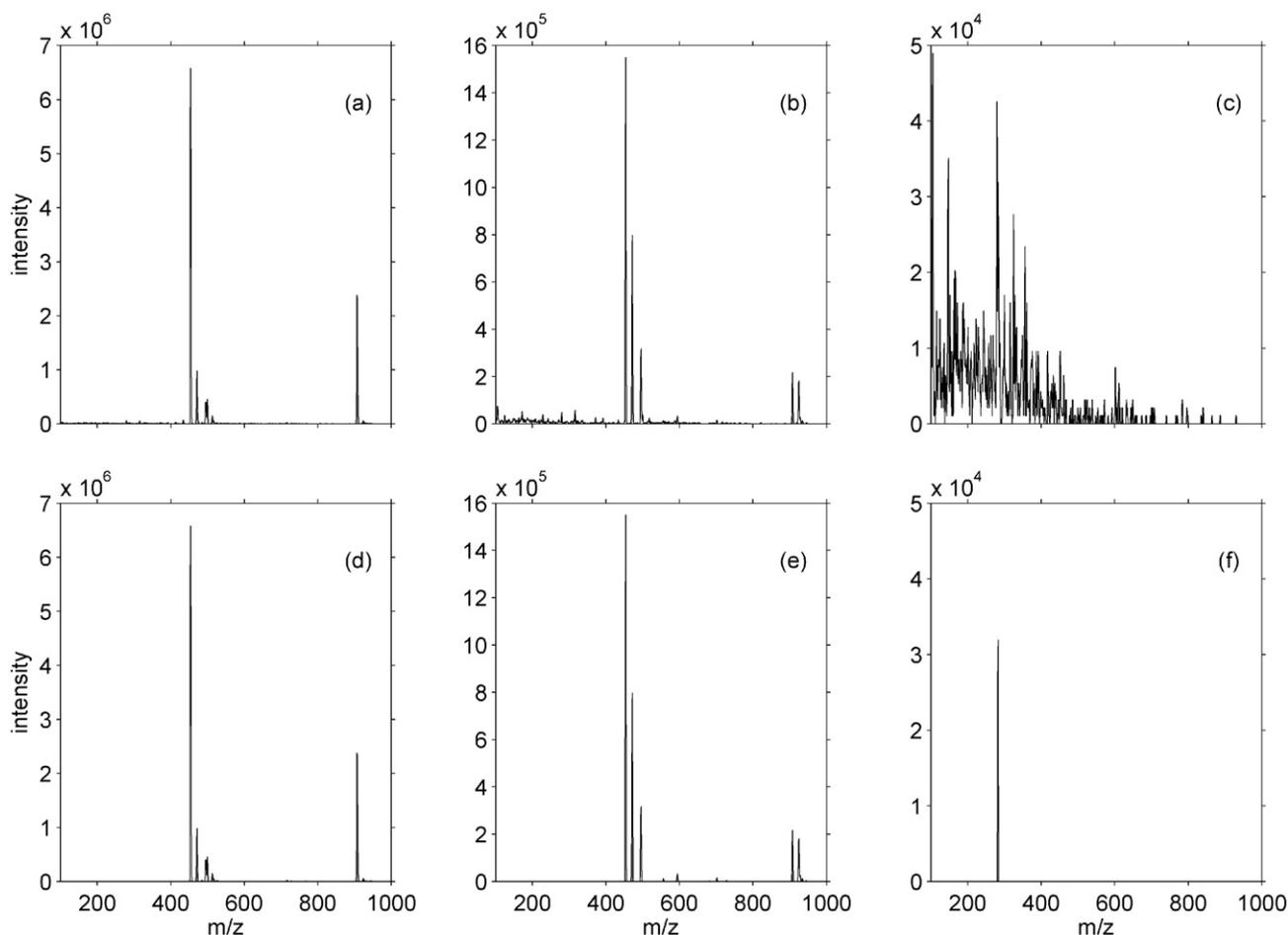
most important parameters, i.e. the retention time and peak width, are automatically estimated with sufficient accuracy by the algorithm. The raw data was quite noisy as can be seen in Fig. 1(a) and (c). After applying algorithms for signal enhancement and peak picking [27] the processed data displayed a significant improvement (shown as gray filled peaks in Fig. 1(b) and (d)). The noise in the peaks is nicely smoothed by the algorithm and mass channels not containing chromatographic information have been removed. A large number of peaks were detected that were not visible in the original TIC. Plotting the raw data TIC or BPC and the sum or max signal of the reconstructed peaks in this way gives a quick indication on how successfully the algorithm has performed. This plot

showed good agreement for the current data sets and the number of detected peaks by the algorithm are found in Table 1. As expected, the number of detected peaks increases when the gradient slope decreases since this facilitates a better separation for most of the components, at least for the current samples. The algorithms alter peak shape and peak area somewhat but if an accurate quantitative analysis is required, the processed data can be converted back to the corresponding  $m/z$  trace prior to integration.

The resulting reconstructed data sets, containing only the detected peaks in the data, were then subjected to the previously developed component tracking algorithm [28], following the line that allows for a simultaneous matching of more than two data

**Table 1**  
The experimental setup and algorithm results.

| Column temp.                             | Mobile phase starting comp. (% ACN) | Mobile phase ending comp. (% ACN) | Gradient time (min) | Number of peaks | Number of initial comp. | Number of automatic matched comp. | Number of erroneous matches |
|--|-------------------------------------|-----------------------------------|---------------------|-----------------|-------------------------|-----------------------------------|-----------------------------|
| 25                                       | 5                                   | 95                                | 12                  | 617             | 31                      | 31                                | 1                           |
| 25                                       | 5                                   | 95                                | 24                  | 655             | 38                      | 31                                | 2                           |
| 25                                       | 5                                   | 95                                | 36                  | 888             | 63                      | 31                                | 3                           |
| 50                                       | 5                                   | 95                                | 12                  | 701             | 59                      | 31                                | 1                           |
| 50                                       | 5                                   | 95                                | 24                  | 736             | 59                      | 31                                | 1                           |
| 50                                       | 5                                   | 95                                | 36                  | 927             | 78                      | 31                                | 1                           |
| Optimized conditions (two step gradient) |                                     |                                   |                     |                 |                         |                                   |                             |
| 50                                       | 14                                  | 41                                | 15                  | 627             | 48                      | 30                                | 0                           |
|  | 41                                  | 95                                | 25                  |                 |                         |                                   |                             |



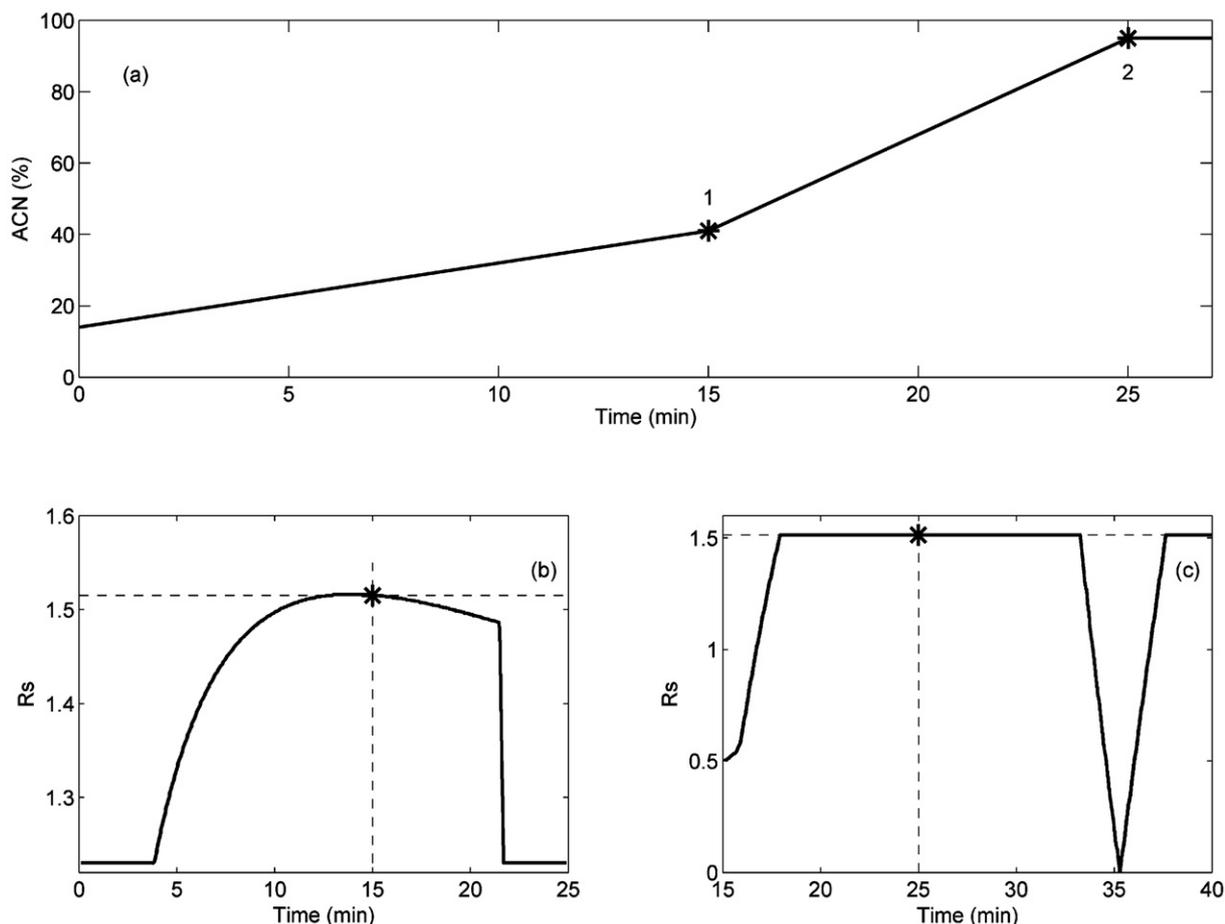
**Fig. 2.** (a) Raw data spectrum of flucinolone acetonide. (b) Raw data spectrum of component #21. (c) Raw data spectrum of component #28. (d–f) Corresponding spectra that were automatically peak detected and found in all six of the data sets prior to optimization as well as in the optimized data set.

sets. In Table 1, the resulting number of detected peaks and initial components are listed. Initial components consist of the chromatographic peaks and the mass spectrum that correspond to a component roughly estimated initially by the component tracking algorithm [28]. These should preferably contain at least one complete component (i.e. the initial components do not need to be pure). If this criterion cannot be fulfilled, and the number of components is over-estimated in some areas, some true components will be separated into two or more fractionated components in the finalized result, usually with one much smaller than the other. These fractionated components are then treated individually, but are always connected in the sense by a common retention time in all data sets where they are detected and tracked, and thus easily spotted by the user, and often does not critically deteriorate the matching abilities.

The number of automatically tracked components that existed in all data sets, extracted from the initial components by the next process in the algorithm, was reported to be 31. The sum of these is shown as the solid line in Fig. 1(b) and (d). Comparing to the gray filled peaks, in the same figures, it is clear that the major part of the data sets could be automatically tracked and only minor components or part of components was unable to be tracked. The smallest tracked component that with certainty was not a fraction of a larger component was 0.08%, whereas the smallest overall was 0.04% of the main component. It is likely that these levels actually are even lower since the signal of the main component is high enough to reach the nonlinear region where the MS detector starts to become saturated.

It can be expected that the peak tracking algorithm make some mistakes. Components with a low S/N ratio or those having many similar neighbouring components can be difficult to evaluate, even manually, and there are cases of inseparable or missing components. Moreover, LC–MS can be rather unpredictable regarding S/N levels, appearance of ghost peaks, and other artefacts between different data sets even though the same sample has been used.

Since only the common mass channels are stored after that the component tracking is performed, the results can be easily visualized for each component and all data sets simultaneously. Any erroneous match is thereby easily spotted. Another approach for finding erroneous matches at this stage is to check the linearity between gradient slope and retention time and search for suspicious behaviour. A third approach is to skip this step and assume that most of the components are matched correctly if a model is to be made from the matches. Verifying the model with a new analytical run will then show incorrect matches deviating in retention time from the model at higher degree compared to the correctly matched ones. Erroneous matches can then be detected and corrected at the cost of some tuning of the instrument and a new analytical run. Manual inspection of the automatic component tracking results showed, however, that one component tracked in all data set was a detected false positive obtained by the sudden change in mobile phase composition in the end of the runs at  $m/z$  112, 132 and 185 (Table 1). This false component could be easily spotted since no change in retention at the different gradient slopes was obtained. Three other erroneous matches were found where either the correct components were missing (i.e. not registered by the algorithm)



**Fig. 3.** (a) The mobile phase gradient predicted for optimal separation of component #21 and #28. (b) Critical resolution between #21, #28 and adjacent peaks as function of gradient time for the gradient time of the first gradient segment (point 1 in (a)). (c) Critical resolution between #2, #28 and adjacent peaks as function of gradient time for the second gradient segment (point 2 in (a)).

because of too low S/N levels, and/or the algorithm selected an alternative candidate component with similar spectrum and intensity. If including all types of matching errors, the automatically proposed matches were thus 95% correct.

A peak list containing the components candidating as a possible match and their measurements of similarity are available after peak tracking [28]. This list can be of assistance for detecting some of the obtained erroneous matches manually. In the case where the true matching component is absent in one data set, the similarity index (i.e. the combined similarity measurements) between the erroneously listed best matching target and candidate component will receive a low number compared to the average score. In the case where the true component exist, but the algorithm picks a component with similar area and spectrum, the true component, often the second best match, will obtain a similarity value close to the top candidate. A top candidate with a low similarity index score or a second best match with a near top score does, however, not necessary mean by default that the top candidate is erroneous, but can indicate that a closer manual examination is needed. Erroneous matches that arise from false components present in all data sets (e.g. fast mobile phase composition shifts on bleeding columns) can generally not be detected by merely using the peak list.

The components left behind by the algorithm that did not pass the criterion to be regarded as a match between the data sets, i.e. the residual components that optimally only consist of data set specific false positives or missing components due to no elution during sampling interval or low S/N levels, can also rather easily be analyzed. A recommended approach is to select the data set that

generated the largest number of residual components and plot the raw data chromatograms for those  $m/z$  ratios present in the current component spectrum. Then repeat this process for all data sets and simultaneously show the result in different windows. The reason for a rejected component can then often be established. None of the residual components had a distinctive estimated area (average = 0.3% of main component) and none could with confidence be manually tracked between all six data sets. One component, also visible in the TIC in Fig. 1(c) shown as a little bump just to the right of the main component is actually a true component having the same spectrum as the main component. In the low temperature data sets, this component coelutes with the main component. Consequently it is practically impossible to separate these components without comparing differences in intensities or area. Thus the algorithm treats the component as missing in the low temperature data sets and is not included in the further considerations. If a component is erroneously placed with the residual components, it can, however, be fruitful to return it in the group of tracked components prior to the following optimization since only components registered as tracked between all data sets will be considered.

### 3.2. The model

The next step is to build a retention and peak width models. In the current application this was made in DryLab, which is a well established simulation program that has been reported to function well in other optimizations with similar samples [4–6,8,13,14,29]. Of course, any optimization program with similar functionality can

**Table 2**  
The main spectral signals and suggested clusters.

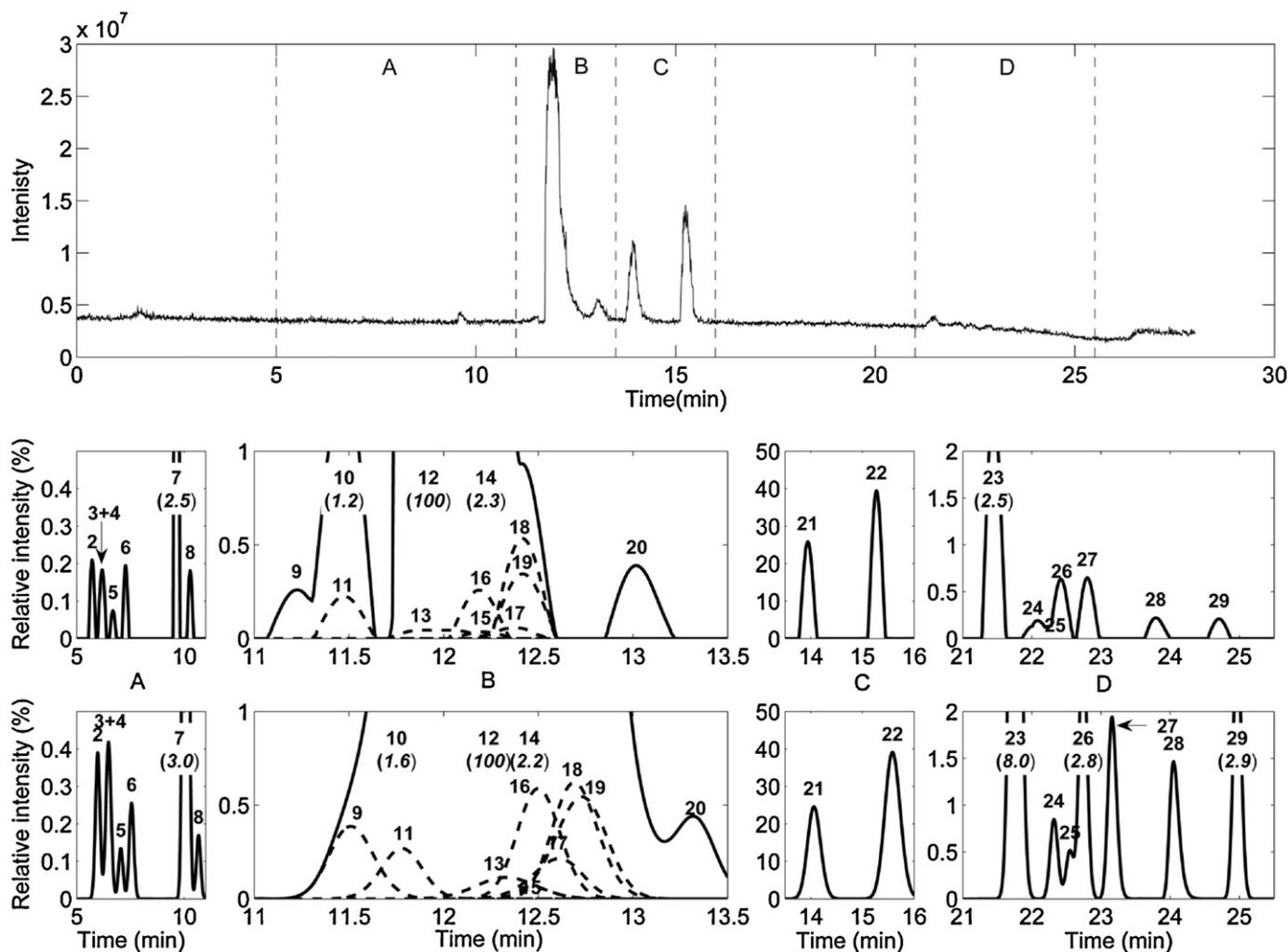
| Component #     | Main signal ( $m/z$ ) | $[M+H-H_2O]^+$ | $[M+NH_4]^+$ | $[M+Na]^+$ | $[M+K]^+$ | $[M+ACN+H]^+$ | $[M+2Na-H]^+$ | $[2M+H]^+$ | $[2M+NH_4]^+$ | $[2M+Na]^+$ | $[2M+K]^+$ | $[2M+2Na-H]^+$ |
|-----------------|-----------------------|----------------|--------------|------------|-----------|---------------|---------------|------------|---------------|-------------|------------|----------------|
| 1 <sup>a</sup>  | 105                   |                |              |            |           |               |               |            |               |             |            |                |
| 2               | 452                   |                |              |            |           |               |               |            |               |             |            |                |
| 3–4             | 498 <sup>c</sup>      |                | x            |            |           |               |               |            |               |             |            |                |
| 5               | 496                   |                | x            |            |           |               |               |            |               |             |            |                |
| 6               | 452                   |                |              |            | x         |               |               |            |               |             |            |                |
| 7               | 434                   |                |              |            |           |               |               |            |               |             |            |                |
| 8               | 468                   | x              | x            |            |           |               |               |            |               |             |            |                |
| 9               | 440                   |                |              |            |           |               |               |            |               |             |            |                |
| 10–11           | 436                   |                |              |            |           |               | x             |            |               |             |            |                |
| 12–13           | 454                   |                | x            |            |           | x             | x             | x          | x             | x           | x          |                |
| 14              | 428                   |                | x            |            |           |               |               |            |               |             |            |                |
| 15–16           | 881                   |                | x            |            |           |               |               |            |               |             |            |                |
| 17              | 490                   |                |              |            |           |               |               |            |               |             |            |                |
| 18–19           | 468                   | x              |              |            |           |               |               |            |               |             |            |                |
| 20              | 434                   |                |              |            |           |               |               |            |               |             |            |                |
| 21 <sup>b</sup> | 454                   |                | x            |            |           | x             |               | x          | x             | x           | x          |                |
| 22              | 434                   |                | x            |            |           | x             | x             | x          | x             | x           | x          | x              |
| 23              | 280                   |                |              |            |           |               | x             |            |               |             |            | x              |
| 24              | 253                   |                |              |            |           |               |               |            |               |             |            |                |
| 25              | 376                   |                |              |            |           |               |               |            |               |             |            |                |
| 26              | 288                   |                | x            |            |           |               | x             |            |               |             |            |                |
| 27              | 634 <sup>d</sup>      |                |              |            |           |               |               |            |               |             |            |                |
| 28 <sup>b</sup> | 283                   |                |              |            |           |               |               |            |               |             |            |                |
| 29              | 299                   |                |              |            |           |               |               |            |               |             |            |                |
| 30 <sup>a</sup> | 392                   |                |              | x          |           |               | x             |            |               | x           |            | x              |

<sup>a</sup> These are not visible in Fig. 4.

<sup>b</sup> These are the optimized components.

<sup>c</sup> This is probably the mass of the  $[M+H-H_2O]^+$  ion because the  $[M+H]^+$  ion ( $m/z$  516) is absent in one of the model data sets but exists in the optimized data (manual inspection).

<sup>d</sup> This is probably the mass of the  $[M+2Na-H]^+$  ion and the  $[M+H]^+$  ion ( $m/z$  589) is absent in one of the model data sets but exists in the optimized data (manual inspection).



**Fig. 4.** (Upper) TIC of the optimized experimental data set, divided into 4 segments marked a, b, c and d. (Middle) The component profiles of the tracked components in the different segments after processing by the component tracking algorithm and (bottom) suggested by Drylab, numbered after elution order. The solid line is the sum of all the component profiles. Dashed lines are some selected individual component profiles. Numbers in brackets are the percentage of the relative intensity of the components with intensity above the visualized range in relation to the intensity of the main component (#12). Numbers without brackets positioned above the solid line indicates that the height of that component is not significantly affected by any of the other components. All numbers with brackets are positioned in time at the peak apex. Please be aware of the different scales in the segment windows. Component number #21 and #28 are the optimized components in the data set.

be used for this purpose. DryLab requires retention times, width, area and the asymmetry of the chromatographic peaks in the data sets to generate models. These values are easily available from the reconstructed component profiles obtained from the component tracking algorithm and can be used with decent results even though they do not perfectly correspond to the peaks obtained from the raw data. For comparison, the actual data elution profiles were estimated by fitting a function described in [30] to the raw data signals corresponding to each component. In general, tailing raw data peaks are less asymmetric in the reconstructed component profiles in comparison with curve fitted data. In the current data set, where many component peaks are tailing, the peak asymmetry factor at 10% height is on average 40% lower and there is not much linearity since the coefficient of correlation,  $r^2$ , is 0.142. This also influences the width (average 14% narrower). The area also becomes lower when comparing to tailing raw data components since a part of the tail is not measured in the reconstructed component profiles, but perhaps more important is that we do have a linear relation. In our case the estimated areas of the reconstructed component profiles was 82% of the curve fitted raw data variant and  $r^2 = 0.977$ . In the current application, retention times are most important and the average retention time differed by only 0.5% when comparing the

reconstructed component profiles with those curve fitted to raw data. It should be emphasized, however, that there is no evaluation of the errors of the fitted curve to the raw data used for the comparison and the results is also highly dependent on the structure of the data sets (i.e. less noisy and less tailing data sets give improved results since both the reconstructed peak detected version and the curve fitted representation will better imitate the actual raw data).

According to DryLab, a decent separation for all detected components would require 98 min at 40 °C. To illustrate the principle it was therefore decided to focus on two components and elute them with an acceptable resolution to adjacent peaks at shortest possible time. One of these components, #21, elutes closely after the main component, has a high intensity and a mass spectrum that is very similar to main component. The other selected component, #28, elutes later and has a much lower intensity. The elution profiles of these are marked in Fig. 1. The relative peak areas were estimated to around 22 and 0.2% compared to the main component. The raw data spectra of these components can be viewed in Fig. 2.

To reduce back pressure and analysis time it was decided to optimize at 50 °C. Models were thus calculated by using data from the three data sets obtained at 50 °C. Optimal conditions were defined as shortest possible analysis time and a resolution to adjacent peaks

of 1.5. As illustrated in Fig. 3 the models suggested that a 25 min gradient employing two linear segments (Table 1) was needed in order to fulfil the requirements.

### 3.3. The optimized data set

An additional experiment was made to confirm the location of the proposed optimal conditions. The TIC of the data set obtained at these conditions can be viewed in Fig. 4 together with the component profiles of the tracked components and the component profile predicted by DryLab. All 30 valid components were correctly matched, whereas the false positive component could not be found in the optimized data set. If reconstructing the optimized data set with the modelled values, Fig. 4, it can be seen that the reconstructed experimental components are eluted earlier compared to the calculated components. Plotting the respective retention times demonstrate that a strong linearity is obtained ( $r^2=0.997$ ), but the experimental data elute in average 14 s earlier than the predicted data. These time differences can be an effect of an alteration of the mobile phase composition due to prolonged time between the model and the verifying experiments. There are also some differences in component width ( $r^2=0.12$ ), asymmetry ( $r^2=0.18$ ) and area ( $r^2=0.998$ ). Better peak width and asymmetry models are expected with data containing better peak shapes. Moreover, the time between generating the models and verifying the optimized run was 4 weeks, consequently the drug might have been somewhat more degraded, even though stored in darkness in a freezer.

The obtained minimum resolution for component #21 and #28 (defined as  $2(t_{RA} - t_{RB})/(w_A + w_B)$ , where  $t_R$  is retention time and  $w$  is the base width) was 2.45 and 2.58, respectively, when using the component profiles for the reconstructed algorithm estimation in Fig. 4(middle) as the result. The actual resolution of the optimized unprocessed data was manually calculated to 2.45 and 2.13. Thus both the processed and actual data gave better resolution compared to the predicted resolution ( $R_s \sim 1.5$ ). Possibly this deviation could be related to a reconfiguration of the LC-MS that took place between the collection of the data used to build the models and the experiment made to verify optimal conditions. Typical deviations between predicted and actual retention time and peak widths are <1% and <15% respectively. The use of 2.1 mm columns and sub 2  $\mu\text{m}$  porous or 2.7  $\mu\text{m}$  superficially porous particles on conventional HPLC equipment will in most cases result in more asymmetric peaks than when using wider columns and/or larger particles. To fully utilize such columns ultra-HPLC equipment is needed. This is related to a larger impact of dead volumes and extra column band broadening in capillaries, fittings and detector cells on peaks with a smaller volume. 2.1 mm i.d. columns are needed for MS analysis but it is likely that broader and more symmetric peaks could have been obtained with larger particles.

In Fig. 2(d–f), the spectra correspond to entities that were automatically found in the optimized data set and also in the six datasets used for modelling. For component #21 and the main component in the sample, the mass ions are plentiful and since also the intensity information is utilized by the algorithm, the components can be discriminated rather easily even though the spectrum is similar for these and also for other components in the neighbourhood. For the much less intense component #28, only two adjacent mass ions with the main signal corresponding to  $m/z$  283 was detected in all data sets. A few numbers of mass ions makes matching more difficult in general, but was successful in this case. Component #28 also had detected peak at  $m/z$  328 (+45) in the higher column temperature data sets but this signal could not be detected in the three data sets sampled at lower column temperatures where the intensities were lower and thus this  $m/z$  are rejected by the algorithm for this component. The reason for different intensities and areas

between the data sampled at a column oven at 25 and 50 °C are probably related to on-column thermal degradation of fluocinolone acetonide [24,31].

Implemented as a part of the peak tracking algorithms is a function for annotation of common adducts, dimers and neutral losses. In Table 2, the main signal (i.e. the  $[M+H]^+$  ion in most cases) is listed together with some suggested clusters that were present and detected in all data sets. By combining the information from the elution profiles, the corresponding mass spectra and the suggested clusters, it could be concluded that some of the tracked components probably are fractionated components as described earlier in Section 3.1, and these were combined in Table 2, which then sums up to 25 components. Noisy data containing a number of asymmetric peaks suffer a greater risk of obtaining fractionated components. A manual fusion of the components in question at an earlier stage (e.g. after tracking the six model data sets) probably could have reduced the number of fractionations when tracking the optimized data. Many of the detected components have the maximum signal close to the  $m/z$  of the main component, suggesting that these really are degradation products of fluocinolone acetonide. It is, however, possible that some have other origins, e.g. the ones with the main signal above  $m/z$  454. These components could be leachables from plastic containers, elastomeric septa or equivalent. None could, however, with confidence be linked to some of the most common types of non-pharmaceutical contaminants [32]. The obtained spectrum could provide indications on what degradation products have been formed, especially when a soft ionization technique such as ESI has been used. However, spiking with known impurities, LC-MS/MS and/or NMR studies are needed for a more comprehensive structure elucidation. Structural elucidation of the impurities and degradation pathways is a natural next step, but is beyond the scope of the current investigation.

## 4. Conclusions

In this paper it has been shown how our recently developed algorithms for highly automated data processing for detection and tracking of chromatographic peaks in noisy LC-MS data can be applied to facilitate LC-MS method development. The algorithms were shown to be successful in the optimization of a separation of components present in a challenging sample containing a large number of components present at low level and with similar or, in some cases identical, mass spectra impeding the possibility to discriminate the components by their spectra alone.

Moreover, despite the relatively low signal to noise ratio and, the in some cases, tailing peaks, the peak tracking algorithm provided accurate results. The algorithms are fast, sensitive and accurate enough to target impurities present at levels relevant for pharmaceutical products.

One important feature of the algorithms is the high degree of automation. Typically the algorithm manages to locate relevant components without any user intervention even for challenging data sets such as the one presented in this paper, where analytes are present at low level and therefore not visible at all in the TIC, BPC or UV chromatograms.

Even though the peak tracking algorithm has primarily been designed for localization and tracking of chromatographic peaks it also improves the quality of the mass spectra. For each component that has been tracked a noise reduced mass spectrum is generated which can facilitate identification and structure elucidation.

The implementation of automated data processing for detection and peak tracking in method development strategies is, in our opinion, likely to not only reduce the time needed for method development but also improve the quality and reliability of the methods produced.

## Acknowledgements

We would like to thank AstraZeneca, Analytical Development, R&D Lund, Sweden and European regional development fund of the European Union for financial support.

## References

- [1] S. Baertschi, *Trends Anal. Chem.* 25 (2006) 758.
- [2] B.A. Olsen, B.C. Castle, D.P. Myers, *Trends Anal. Chem.* 25 (2006) 796.
- [3] G. Chen, B.N. Pramanik, Y. Liu, U.A. Mirza, *J. Mass Spectrom.* 42 (2007) 279.
- [4] I. Popovic, D. Ivanovic, M. Medenica, A. Malenovic, B. Jancic-Stojanovic, *Chromatographia* 67 (2008) 449.
- [5] M. Pfeffer, H. Windth, *Fresenius J. Anal. Chem.* 369 (2001) 36.
- [6] I. Molnár, H. Rieger, K.E. Monks, *J. Chromatogr. A* 1217 (2010) 3193.
- [7] W. Li, H.T. Rasmussen, *J. Chromatogr. A* 1016 (2003) 165.
- [8] R.M. Krisko, K. McLaughlin, M.J. Koenigbauer, C.E. Lunte, *J. Chromatogr. A* 1122 (2006) 186.
- [9] R.T.I. Smith, S. Guhan, K. Taksen, M. Vavra, D. Myers, M.T.W. Hearn, *J. Chromatogr. A* 972 (2002) 27.
- [10] E.F. Hewitt, P. Lukulay, S. Galushko, *J. Chromatogr. A* 1107 (2006) 79.
- [11] M. Dumarey, R. Sneyers, W. Janssens, I. Somers, Y. Vander Heyden, *Anal. Chim. Acta* 656 (2009) 85.
- [12] X. Cui, B. Shao, R. Zhao, Y. Yang, J. Hu, X. Tu, *Rapid Commun. Mass Spectrom.* 20 (2006) 2355.
- [13] C.C. Corredor, J.A. Castoro, J. Young, *J. Pharm. Innov.* 4 (2009) 121.
- [14] I. Molnar, *J. Chromatogr. A* 965 (2002) 175.
- [15] <http://www.molnar-institut.com>, accessed 06.10.2010.
- [16] <http://www.acdlabs.com>, accessed 06.10.2010.
- [17] <http://www.chromsword.de>, accessed 06.10.2010.
- [18] <http://www.datalys.net>, accessed 06.10.2010.
- [19] M. Moberg, J. Bergquist, D. Bylund, *J. Mass Spectrom.* 41 (2006) 1334.
- [20] M. Moberg, K.E. Markides, D. Bylund, *J. Mass Spectrom.* 40 (2005) 317.
- [21] M. Moberg, D. Bylund, R. Danielsson, K. Markides, *Analyst* 125 (2000) 1970.
- [22] M.R. Euerby, P. Petersson, *J. Chromatogr. A* 994 (2003) 13.
- [23] N. Lundell, K. Markides, *J. Chromatogr.* 639 (1993) 117.
- [24] T. Teutenberg, *Anal. Chim. Acta* 643 (2009) 1.
- [25] G. Miolo, S. Caffieri, D. Dalzoppo, A. Ricci, E. Fasani, A. Albin, *Photochem. Photobiol.* 81 (2005) 291.
- [26] A. Ricci, E. Fasani, M. Mella, A. Albin, *J. Org. Chem.* 68 (2003) 4361.
- [27] M. Fredriksson, P. Petersson, B. Axelsson, D. Bylund, *J. Sep. Sci.* 32 (2009) 3906.
- [28] M.J. Fredriksson, P. Petersson, B. Axelsson, D. Bylund, *J. Chromatogr. A* 1217 (2010) 8195.
- [29] C. Liu, P. Zhu, M. Liu, *J. Chromatogr. A* 857 (1999) 167.
- [30] J.R. Torres-Lapasio, J.J. Baeza-Baeza, M.C. Garcia-Alvarez-Coque, *Anal. Chem.* 69 (1997) 3822.
- [31] R.A. Kenley, M.O. Lee, L. Sukumar, M.F. Powell, *Pharm. Res.* 4 (1987) 342.
- [32] B.O. Keller, J. Sui, A. Young, R.M. Whittall, *Anal. Chim. Acta* 627 (2008) 71.